DOCUMENT RESUME

ED 304 095 IR 013 631

AUTHOR Anbar, Michael
TITLE Cue-Free Computerized Interactive Tests--Computer Emulation of Oral Examinations.
PUB DATE May 88
NOTE 6p.; Paper presented at the American Association for Medical Systems and Informatics Congress (San Francisco, CA, May 5-7, 1988).
PUB TYPE Viewpoints (120) -- Reports - Descriptive (141) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Authoring Aids (Programing); *Computer Assisted Testing; *Computer Simulation; Higher Education; *Medical Schools; *Microcomputers; Objectives; Tests
IDENTIFIERS Medical Students; *Natural Language; *Oral Examinations; State University of New York Buffalo

ABSTRACT

This discussion of the use of microcomputer software in a medical school class for the purposes of emulating oral examinations begins by stating the three major goals of testing in medical school. The limitations of multiple choice tests and oral examinations are then discussed, and the use and suitability of computers to administer and to avoid the shortcomings of these kinds of tests are described. The advantages and disadvantages of using microcomputers to administer short answer and essay type examinations are also considered. The paper next describes the authoring software system, Computer Administered Socratic Instruction Program (CASIP), which is designed to create computer assisted instruction that responds to natural language input. The Computer Administered Socratic Examination Program (CASEP), a modified version of CASIP which was developed to administer cue-free interactive computerized tests run on IBM PC or compatible personal computers, is described in more detail. Steps in authoring a quiz using CASEP are outlined, tests of this method of quiz administration in a sophomore course on medical technology are briefly discussed, and the advantages of this methodology are summarized. (7 references) (CGD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

PROCEEDINGS OF AAMSI CONGRESS 1988
San Francisco, CA 5/5-7/88

CUE-FREE COMPUTERIZED INTERACTIVE TESTS - COMPUTER EMULATION OF ORAL EXAMINATIONS

Michael Anbar PhD

School of Medicine and Biomedical Sciences,
State University of New York, Buffalo, NY, 14214

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Michael Anbar
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED304095

IR013631

Oral examination is the most effective method to test for reasoning and decision making skills in medical practice. This approach has, however, severe drawbacks; paramount among these is the faculty time it necessitates. This and several other limitations of oral examination can be alleviated by using the computer to emulate oral testing. In this paper we describe recently developed software which makes this possible. We report here our experiences with the use of this software using microcomputers in a medical school class.

Testing in medical school, as in most other academic settings, has three major goals:

A. To determine the extent that the instructor's teaching objectives were met, by assessing the students' level of knowledge and their ability to use this knowledge.

B. To provide the students with feedback as to how well they have achieved the goals of the course.

C. To differentially evaluate the individual performances of students, reward the high achievers, and identify those who need coaching or help.

When used for licensing, i.e., quality control of physicians, testing is aimed at identifying those who do not meet minimal standards of knowledge and skills.

These goals plus the fact that tests motivate students to study, thus "keeping them on their toes," justify the substantial efforts invested in testing in medical schools.

Ideally, we would like to talk with students individually after each class to find out what they have learned, i.e., to what extent have we achieved our instructional goals. Practically, we may, at best, do this just once or twice a semester. Because of time limitations we give up on personal interaction altogether and administer written tests of one type or another even during clinical clerkships. The size of medical classes even limits the amount of time available for evaluating each written test, thus we resort to multiple choice exams, to facilitate automatic computerized grading.

Multiple choice tests limit the scope and number of optional solutions to a given question. Most problems in medicine have more potential alternative solutions than are available in even the most comprehensive multiple choice

BEST COPY AVAILABLE

test. Furthermore, there is no way to find out if the "correct" answer was given for the right reason, by chance, or for a wrong reason. It is also impossible to know whether a "wrong" answer was given for a valid and justifiable reason. Moreover, even ingeniously authored multiple choice tests cannot ascertain whether the student could have come up with the right answer without cues. Finally, there is no way to determine if students have actually mastered the subject matter, or have just learned to identify distractors. Consequently, students trained to respond exclusively in the multiple choice environment may be inadequately trained to cope with real life clinical situations,(1) and therefore rarely are multiple choice tests used in qualifying examinations for a license, where it is critical to unambiguously measure the competence of the candidate.

Computers are very well suited to administer multiple choice tests. Random selection of questions from a pool, making subsequent questions dependent on previous answers and automatic scoring can make this testing modality more efficient than it ever was. However, non of these features alleviates the major limitations of this testing modality.

Microcomputers can also be used to administer essay type exams. Using the computer to input short essay answers has several advantages. A legible printout of the student answers is produced, which eliminates intentional or unintentional obfuscation of crucial words. "Blind" grading of answers is facilitated because of the ease of creating multiple copies of exam answers to be distributed to different graders, and the ease with which answers can be graded question by question instead of sequentially grading all the answers of each student. The latter modality facilitates a more standardized differential evaluation and avoids biasing the grade for a given answer by the student's performance on previous ones.

Computer based short answer and essay-type tests suffer, however, from many of the same problems as hand-written tests of this type. They are time consuming to grade, and do not insure that the student will address exactly the expected points or provide the information desired. It is for these reasons and others that the crucial part of medical specialty board examinations is administered orally, in spite of the substantial cost involved. When it comes to accurately assessing the students' knowledge and their ability to cope with a real life situation, the oral examination is still the best tool available.

Oral exams, however, are expensive and difficult to standardize. They require examiners with a mastery of the knowledge to be tested and who are able to spend sufficient time with each student. The examiners must also be immune to subtle cuing and bias on their part that can lead to inaccurate grading and to different grading of similar performances. Moreover, oral examinations are often biased by personality differences between the examiner and the examinee.

The microcomputer, which can improve the use of tests based on multiple choice and essay questions, could be used to emulate the testing environment of an oral examination, thus avoiding their shortcomings. This might best achieve all three goals of testing without increasing the overall time commitment of the faculty. Microcomputers can affordably "spend time" with students individually. They can also avoid the potential bias and cuing associated with person to person interaction.

To fulfill this function the computer must be capable of "understanding" answers given in unrestricted natural language. Once an answer is precisely "understood", or recognized by the computer, the test should be capable of branching, and present follow-up questions based on the answer given. The test should be composed of several question-answer cycles, each of which probes in depth the student's knowledge in a certain area and assures that the student has mastered the knowledge tested or has the decision-making ability to handle a given situation in a cue free environment. The test should be programmed to

score each answer and generate a record that can be handled by a database, and at the same time record verbatim the answers given during the test. The former records should enable the instructor to obtain an objective grade, while the latter facilitate a careful analysis of the student's reasoning when the score is low, to prevent penalizing a student for choosing an unconventional but justifiable solution. All these requirement have been recently met.

We have developed a system that allows us to administer cue-free interactive computerized tests run on the IBM PC or PC clones. The software is a modified version of CASIP (Computer Assisted Socratic Instruction Program), a driver program written in the C language.(2,3) CASIP is comprised of an advanced parser (that facilitates the recognition of words and phrases in context) and a conditional brancher that controls the response of the program according to identified answers of the student. CASIP, which is being used at our school to author instructional units and clinical case simulations in particular, has been described before,(3) as has the rationale that led to its development.(4) Essentially, CASIP facilitates authoring of cue-free CAI (with natural language input) using any word processor and without prior knowledge of computer programming.

CASIP is not the only medical CAI program that responds to natural language input -- the same didactic needs were identified by others. They have led to the independent development of a similar program, also written in the C language, at the Medical College of Georgia,(5,6) and recently of another program, written in PROLOG, at the School of Medicine, University of Missouri.(7) In brief, it has been independently shown at least in three medical schools that natural language computer aided instruction is feasible. We have now extended the use of this type of software to fulfill another important function in medical education -- that of interactive testing.

CASEP (for Computer Administered Socratic Examination Program) differs from CASIP by allowing random selection of question-answer cycles from a pool (each cycle can contain several levels of questions in depth, or a sequence of decisions, concerning a certain medical problem), by terminating the session after a prescribed number of question-answer cycles have been completed, and by generating three kinds of permanent record files: a verbatim record of each question-answer-comment interchange during the session, a record of the student's score profile (providing a separate score for each cycle) plus the total score, and a record of the program parameters used during the sessions. From the latter record the instructor can derive the branching pattern followed by the program during the session. The latter two records are written to "hidden" files inaccessible to the students.

The student can obtain a printout of the first file. This provides additional useful feedback for the student and for the interested instructor. The second file can be read by another computer program to incorporate its contents into the class record. The third record is used by the instructor during trial runs of the session, when the author-instructor has the opportunity to streamline the "oral" examination.

What does it take to author a quiz using CASEP? First, one has to identify "issue" questions such as "Of all computer applications in medicine which are the least cost effective?", "What diagnostic technique could you use to determine whether your patient has an occlusion of his right external carotid artery?", or "What can you tell a patient who refuses a CT of the head following a concussion, on the grounds that she is afraid of radiation?" These questions are of the kind asked in our [illegible] "Clinical Biostatistics", which covers the fundamentals and applications of medical testing. What these questions have in common is that non of them has a clear-cut answer. Each of them requires recall of several options and the making of choices among them. The author must anticipate the possible answers to each question and think of the characteristic words that might appear in the answers.

BEST COPY AVAILABLE

"Characteristic words" here means words which would differentiate a given answer from any of the alternative answers. Sometimes the presence of a single word makes an answer distinctive, whereas in other cases one may need three or four words in a prescribed sequence to assure unambiguous interpretation of an answer. CASEP like CASIP allows one to prescribe the recognition of the presence or absence of any desirable number of words in a sentence. It also allows one to specify the number of unrecognized words between each of the recognized words. Once the characteristic words have been identified, the author is expected to specify logical synonyms (in the context of the anticipated answer) for each of the characteristic words. CASEP allows the recognition of any number of logical synonyms in lieu of a characteristic word.

The next step for the author is to instruct the program to branch and display the appropriate comment and follow-up question when an answer has been recognized. At this point the author must anticipate answers to the second level question and identify them by their characteristic words. This process may be repeated as many times as prescribed by the desirable level of the probing process. There is also the option of accepting any answer at a certain stage of probing and moving on to the next question on the topic or to another randomly selected question-answer cycle. The depth of probing may differ for different primary answers. It may range from zero, where a recognized primary answer triggers a single response before moving to the next (randomly chosen) question-answer cycle, up to five or more levels of sequential interactive questions.

After all the potential answers of the different options of a given question-answer cycle have been unambiguously recognized to a desirable depth, the question-answer cycle is ready for preliminary testing. The quiz is given to several students or faculty in a pilot test to identify primary or secondary answers that were not anticipated by the author. New legitimate answers are then included among the corresponding primary or secondary answers. The quiz is then tested and corrected again until the author is satisfied, based on the recorded performance, that the quiz recognizes satisfactorily over 95% of the answers that were given by different people and there are no errors of ambiguity of recognition. There always remains the possibility that a certain answer of a student will not be recognized. In this case the computer will ask the student to repeat that answer in other words. When the quiz reaches this level of reliability it is ready for use in class.

The authoring effort required to produce a typical question-answer cycle of a quiz is about two hours. To this one must add another hour of testing and correcting, plus another hour of revising the program following two or three pilot tests. In all, a quiz comprising six question-answer cycles requires about 24 hours of authoring effort. This effort may look excessive when compared with that necessary to generate a multiple choice quiz, but it is comparable to the effort of correcting 6 short essay answers of 120 students (assuming that it takes just 2 minutes to read, comment on and grade each answer).

Remembering that the same computerized quiz can be used year after year in the same class, or in different schools in parallel, makes the interactive open ended quiz cost-effective, in spite of the substantial authoring effort. It must also be realized that this type of testing, unlike multiple choice quizzes or even short essays, is not invalidated if students are aware of the nature of the questions beforehand, since their grades depend primarily on the way they justify their answers. Therefore, one may effectively use quizzes comprised of randomly selected subsets of question-answer cycles from a pool of questions that may be known to the class ahead of time. If this pool is sufficiently large, one may even use the same program as an excercise.

We have tested this method of quiz administration on several groups of 16 students in our sophomore course on medical technology. This was done in

parallel with a computer-based traditional non-interactive quiz comprised of short essay questions on the same material. A comparison between the performance of the students on the two types of quizzes, using rank order statistics, showed a high level of correlation. These highly encouraging results will be presented at the conference.

In summary: A new computer-based testing methodology that essentially emulates the classical oral examination method has been developed and tested. We have shown that computers can be used to test knowledge in depth and thus avoid the pitfalls of multiple choice testing, without paying an unacceptable price in a setting of mass education. In fact, the cost effectiveness of our testing method increases with the size of the class or the number of classes that utilize the same test. In view of its didactic advantages and the higher precision of knowledge assessment offered by this cue-free computer-based method, it may become a method of choice for testing knowledge or assessing decision making-skills.

REFERENCES

1. Anbar M. Open-ended computer based medical instruction - avoiding the pitfalls of multiple choice behavior. Resident & Staff Physician 1986; 32: 91-100.

2. Anbar M. CASIP - A novel authoring tool for open-ended natural language CAI. Proc. ADCIS 28th International Conf. 1986: 1-4.

3. Anbar M. CAI -- A way to avoid the pitfalls of multiple-choice behavior in medical practice. Medical Electronics 1987; 18: 118-124.

4. Anbar M. Socratic CAI - philosophy and experience. Proc. AAMSI Sixth Annual Joint Natl. Congress, Levy A.H., Williams B.T. (eds.) 1987:286-288.

5. Abdulla A.M., Watkins L.O., Henke J.S. The use of natural language entry and laser videodisk technology in CAI, J Med Educ 1984; 59: 739-745.

6. Abdulla A.M., Henke J.S., Watkins L.O. Computer-aided learning: experiences, perspectives and promises. J Am Coll Cardiol 1987; 9:678-683.

7. Sylvester J.M., Gardner D.W., A natural language interpreter for reading insulin orders. Proc. AAMSI Sixth Annual Joint National Congress, Levy A.H., Williams B.T. (Eds.) 1987: 283-285.